SCIENTIFIC REPORTS

OPEN

Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach

Received: 15 May 2017
Accepted: 6 September 2017
Published online: 29 September 2017

H. A. Gaspar[1,2] & G. Breen[1,2]

Using successful genome-wide association results in psychiatry for drug repurposing is an ongoing challenge. Databases collecting drug targets and gene annotations are growing and can be harnessed to shed a new light on psychiatric disorders. We used genome-wide association study (GWAS) summary statistics from the Psychiatric Genetics Consortium (PGC) Schizophrenia working group to build a drug repositioning model for schizophrenia. As sample size increases, schizophrenia GWAS results show increasing enrichment for known antipsychotic drugs, selective calcium channel blockers, and antiepileptics. Each of these therapeutical classes targets different gene subnetworks. We identify 123 Bonferroni-significant druggable genes outside the MHC, and 128 FDR-significant biological pathways related to neurons, synapses, genic intolerance, membrane transport, epilepsy, and mental disorders. These results suggest that, in schizophrenia, current well-powered GWAS results can reliably detect known schizophrenia drugs and thus may hold considerable potential for the identification of new therapeutic leads. Moreover, antiepileptics and calcium channel blockers may provide repurposing opportunities. This study also reveals significant pathways in schizophrenia that were not identified previously, and provides a workflow for pathway analysis and drug repurposing using GWAS results.

Genome-wide association studies (GWAS) have been performed on numerous human disorders and traits[1], uncovering thousands of associations between disorders or quantitative phenotypes and common genetic variants, usually single nucleotide polymorphisms (SNPs), that 'tag' or identify specific genetic loci. Summary statistics from hundreds of GWASs are freely available online, including those from the Psychiatric Genetics Consortium (PGC) Schizophrenia working group. Schizophrenia is a complex disorder with a lifetime prevalence of ~1%, significant environmental risk factors, and a heritability of 65–85%[2] that has been suggested to be highly polygenic in nature[3]. As with other complex genetic disorders, the application of GWAS to schizophrenia has identified multiple disease susceptibility loci. In 2014, over 100 robustly associated loci were identified in a GWAS meta-analysis by the PGC[4]. Similar progress is underway in other psychiatric disorders, with new GWAS reports expected for attention deficit hyperactivity disorder, autism, major depressive disorder, anorexia nervosa, and bipolar disorder in the next year. However, a key question arises: how can the emergence of new and well powered GWAS data inform the development of new therapeutics?

Most attention on the therapeutic utility of GWAS has focused on the identification of individual drug targets[5]. Nelson *et al.* recently demonstrated that the proportion of drug mechanisms with genetic support increases from 2.0% at the preclinical stage to 8.2% after successful approval[6]. Results from genetic studies can also guide repurposing - the finding of new indications for known drugs[7–9]. Recent studies have also shown how pathway analysis on GWAS data could help discover new drugs for schizophrenia[10–12]. However, these studies, as well as studies focused on single genes or targets, have generally lacked a step to show if a GWAS has sufficient power to reliably identify known drugs; this is a critical step that would lend confidence to the discovery of novel drug associations in GWAS data.

[1]King's College London, Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, London, UK. [2]National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, London, UK. Correspondence and requests for materials should be addressed to H.A.G. (email: helena.gaspar@kcl.ac.uk)

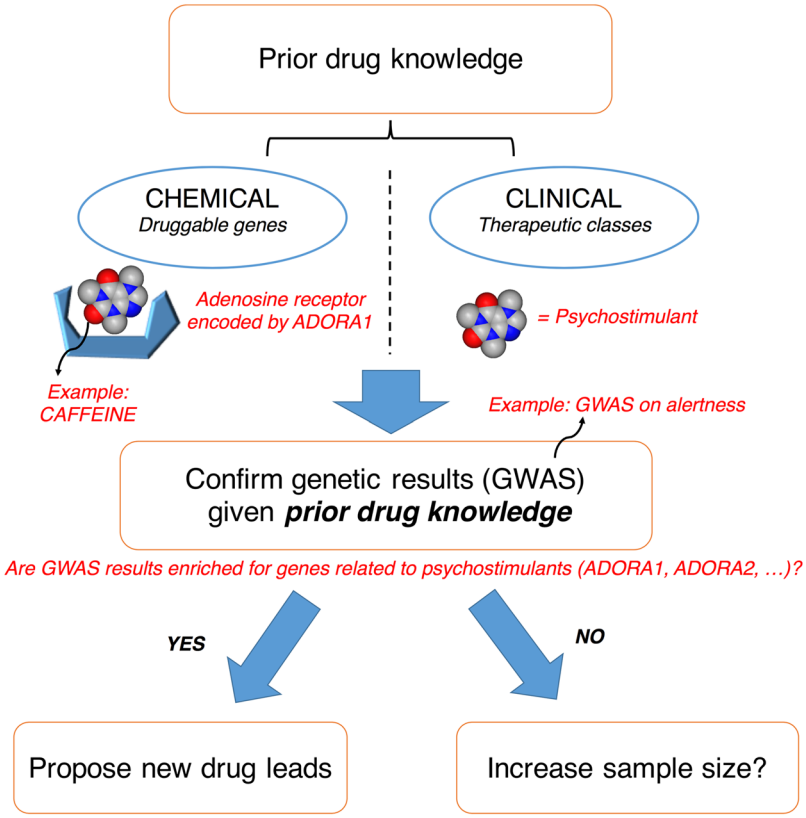

Figure 1. Using drug knowledge to corroborate genetic results. Drug knowledge, encompassing therapeutic classes and druggable genes (*e.g.*, caffeine is a psychostimulant targeting adenosine receptors), may be used to confirm the ability of a GWAS to find known drugs for a given trait (*e.g.*, alertness). Novel targets and potential drugs could then be found in genetic results.

Mining of data available on drug-gene interactions (Fig. 1) allows the combination of individual drug targets into "drug pathways" represented by sets of genes that encode all targets of a given drug or potential novel therapeutic. Any drug can be represented by such a gene-set derived from its drug activity profile, and assigned a p-value generated by pathway analysis assessing the association of a given drug gene-set with the phenotype. An enrichment curve can be drawn for any particular group of drugs using the entire dataset of drugs ranked by p-value. The associated area under the enrichment curve (AUC) provides a simple way to assess the enrichment of any class of drug for a specific disorder. To corroborate a drug repurposing model, we propose to test the enrichment of a known class of drugs, such as antipsychotics for schizophrenia and anxiolytics for anxiety disorders.

In this article, we performed pathway analysis to assess the significance of drugs in schizophrenia GWAS. We analysed and compared three successively larger schizophrenia studies from the PGC Schizophrenia working group: SCZ-PGC1[13], SCZ-PGC1+SWE[14], and SCZ-PGC2[4]. We also analysed the complete SCZ-PGC2 GWAS for the associations of gene families, gene ontology (GO) pathways, canonical pathways, disease pathways, drugs and drug classes with schizophrenia. A common problem in pathway analysis is the interpretation of the top pathways. We propose a new workflow to visualise and cluster significant biological pathways by accounting for pathway similarities as well as pathway significance, based on a kernel variant of the Generative Topographic Mapping approach[15,16].

Results

An analysis of druggable genes was conducted using SCZ-PGC2, excluding the extended major histocompatibility complex (chr6:25652464-33771788). The total number of druggable genes with data in SCZ-PGC2 was 4298. A druggable gene Manhattan plot is presented in Fig. S1a in Supplement 1. We applied two Bonferroni cut-offs: one for the druggable genome (0.05/4298 = 1.163e-5), and one for the whole protein-coding genome (0.05/19870 = 2.516e-6). All druggable genes satisfying the druggable genome cut-off were considered significant, for a total of 123 significant druggable genes with experimentally characterized proteins, excluding the MHC (cf. Table S11 in Supplement 2), divided into druggability Tiers indicating the corresponding target druggability level (cf. Methods). 100 genes were below the protein-coding threshold: 38 Tier 1 genes, 12 Tier 2 genes, 25 Tier 3 A genes, and 24 Tier 3B genes; another 24 genes were below the druggable genome threshold: 8 Tier 1 genes, 8 Tier 2 genes, 5 Tier 3 A genes, and 3 Tier 3.

Calcium voltage-gated channel subunits (*CACNA1I*, *CACNA1C* and *CACNB2*) and several targets of neurotransmitters were significant (Fig. S2 in Supplement 1): cholinergic receptors (the cluster of genes *CHRNA3-CHRNA5-CHRNB4* and *CHRM4*), dopamine receptor D2 (*DRD2*), glutamate metabotropic receptor

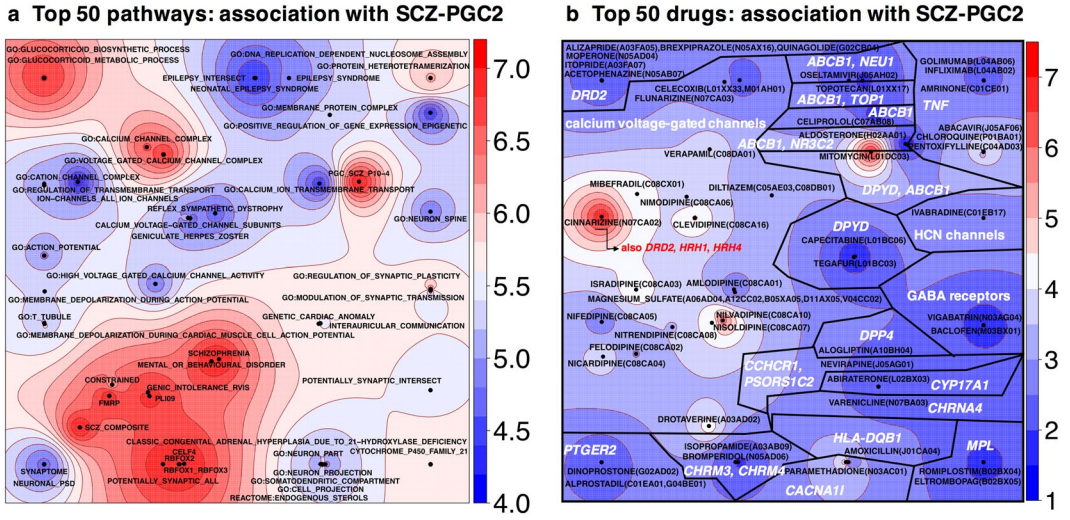

Figure 2. Pathway maps obtained using k-GTM (kernel Generative Topographic Mapping), a dimension reduction algorithm which projects pathways onto a 2D map. The points are gene-sets, positioned according to gene composition. The map is colored by $-\log_{10}(p)$, which measures the degree of association of a gene-set with schizophrenia. (**a**) Top 50 pathways in schizophrenia SCZ-PGC2 GWAS: GO ontology, canonical pathways, gene families or disease gene-sets defined in the Open Targets Platform. All these pathways are FDR-significant according to Benjamini and Hochberg's q-values, whereas only 28 are Bonferroni-significant. (**b**) Top 50 drugs with identified ATC codes, with target information mined from DGIdb and K_i DB. Labels indicate the ATC code for each drug, such as N05 for psycholeptics, as well as the most significant gene(s) in each segment of the map.

3 and glutamate ionotropic receptor NMDA type subunit 2A (*GRM3* and *GRIN2A*), gamma-aminobutyric acid type B receptor subunit 2 (*GABBR2*) and opioid receptor delta 1 (*OPRD1*).

The significant druggable genes were investigated for an overlap with the significant schizophrenia (SCZ) loci (cf. Table S11 in Supplement 2). With a 35 kb upstream, 10 kb downstream window to include regulatory regions, 73 of these genes overlapped with significant SCZ loci, and expanding to a 500kb-500kb window to observe LD (linkage disequilibrium) patterns, 10 other genes were in LD with these loci. Only 40 genes remained independent from SCZ loci: these genes do not contain genome-wide significant SNPs but several SNPs which are suggestively significant. This is the case, for example, for *GABBR2*, *OPRD1* and *NOS1* (Fig. S3 in Supplement 1).

A STRING[17] PPI (protein-protein interaction) network of the 123 top genes was created to identify hub genes; this network is highly connected, with 721 interactions against 454 expected (Fig. S4 in Supplement 1). Normalized betweenness and node degree were computed for each of the 123 genes to investigate their connectivity inside a network only formed with the 123 genes or with 498 genes including the 123 and all significant protein-coding genes outside the MHC (cf. Table 11 in Supplement 2).

Amongst Tier 1 targets (best potential druggable targets), top genes with normalized degree $>5\%$ are: *CACNA1I*, *CHRM4*, *CHRNA3*, *CHRNB3*, *MMP16*, *OPRD1*; on the other hand, Tier 1 hub genes with normalized betweenness $>5\%$ are *MAPK3* and *F2*, and $>2.5\%$: *AKT3*, *DPYD*, *TLR9*, *FGFR1*, *MARK2*, *NOS1*. Recorded mouse studies for hub genes with effect on behaviour or the nervous system are given in Table 12 in Supplement 2, including *CACNA1I*, *CHRM4*, *CHRNA3*, *CLCN3*, *CNTN4*, *NEK1*, *OPRD1*, *F2*, *MAPK3*, *POMC*, *ATP2A2*, *FGFR1*, *FURIN*, *EP300*, *CLU*, *MARK2* and *NOS1*.

We built pathway maps colored by association with SCZ-PGC2 using the kernel generative topographic mapping approach (k-GTM) to identify significant gene-sets with similar gene content (including MHC). The top 50 biological pathways from SCZ-PGC2 pathway analysis were thus mapped onto a 2D map colored by association with schizophrenia in Fig. 2a; the top 50 drugs with identified ATC (Anatomical Therapeutic Chemical) codes were mapped onto another map in Fig. 2b (associated data in Tables S7 and S8 in Supplement 2). Out of the 13,572 biological pathways, 28 reach Bonferroni significance and 112 have FDR q-value $<5\%$. For drugs with ATC code, only five are Bonferroni-significant, and 13 have FDR q-value $<5\%$. Among enriched biological pathways, we find genic intolerance, mental disorders, synapse and neuron pathways, pathways related to histones and nucleosomes, transmembrane transport and ion channels, and epilepsy pathways (Fig. 2a). Top drugs on the map are mainly driven by calcium voltage-gated receptors, *DRD2*, acetylcholine receptors, GABA receptors, HCN channels, or some other individual genes (Bonferroni-significant: *DPYD*, *DPP4*, *CCHCR1*, *PSORS1C2*, *CYP17A1*, *MPL*, *NEU1*, *MPL*, *TNF*, *HLA-DQB1*, *ABCB1*). The top FDR-significant drugs targeting calcium channels are cinnarizine, nilvadipine, parmethadione, clevidipine, isradipine, mibefradil, drotaverine, nisoldipine, verapamil, nicardipine and nimodipine.

The enrichment of ATC drug classes in the latest schizophrenia GWAS (SCZ-PGC2) is reported in Fig. 3a. The enrichment is assessed using the AUC, where AUC = 100% indicates optimal enrichment and AUC = 50% a random result. AUC p-values were computed using Wilcoxon-Mann-Whitney's test and a Bonferroni threshold (1.10^{-3}) was applied to identify significant drug classes, accounting for 49 tests. Antipsychotics (AUC = 75%,

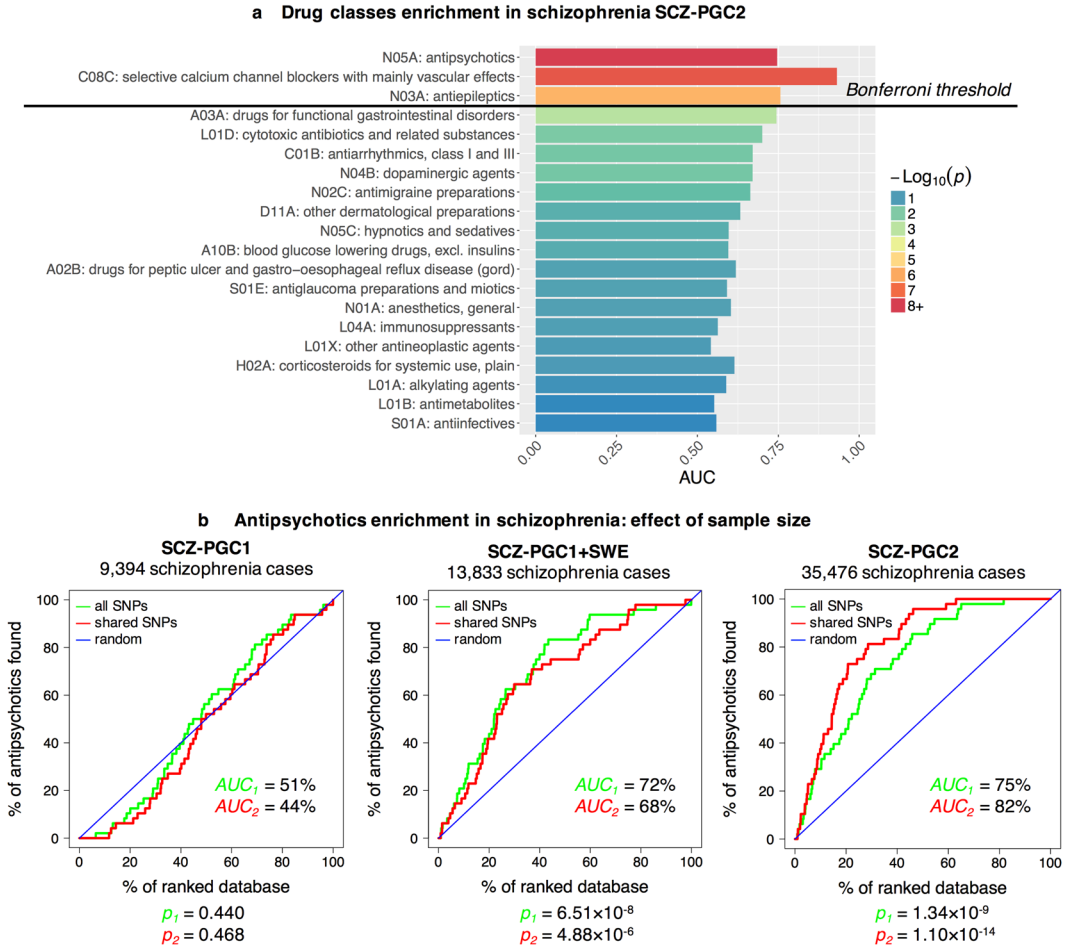

Figure 3. (**a**) Enrichment of top ATC drug classes in SCZ-PGC2 GWAS. AUC is the area under the enrichment curve, and p-values are derived from Wilcoxon-Mann-Whitney's test, which assesses whether drugs of a given class have a higher association with schizophrenia than expected by chance. **(b)** Antipsychotic enrichment in schizophrenia GWASs as a function of sample size. The figure shows enrichment curves for antipsychotics (ATC code N05A), using three GWASs with increasing sample sizes. The expected "random" enrichment curve is indicated in blue. The red enrichment curve is based on SNPs shared between the three studies, and the green enrichment curve uses all SNPs available in a study. Corresponding areas under the curve (AUC) and p-values (*p*) are provided.

$p = 1.342 \times 10^{-9}$), selective calcium channel blockers with mainly vascular effects (AUC = 93%, $p = 3.427 \times 10^{-8}$), and antiepileptics (AUC = 76%, $p = 1.814 \times 10^{-6}$) were significant.

Antipsychotics enrichment curves were generated for SCZ-PGC1, SCZ-PGC1 + SWE and SCZ-PGC2 (Fig. 3b), using only SNPs present in all three studies ("shared SNPs") or all SNPs available in each study. The p-values associated to the AUC were not corrected for multiple testing, since only three planned comparisons were made. For SCZ-PGC1, the antipsychotics enrichment is equal to a random result (with shared SNPs: $AUC = 44\%$, $p = 0.468$); the enrichment is moderate for SCZ-PGC1+SWE (68%, $p = 4.88 \times 10^{-6}$), and high (82%, $p = 1.10 \times 10^{-14}$) for SCZ-PGC2. As the sample size used in schizophrenia GWAS increases (and consequently the statistical power), so does the enrichment for antipsychotics.

The proteins targeted by drug classes enriched in SCZ-PGC2 were investigated using PPI networks (Fig. 4a). The analysis revealed that enriched drug classes (selective calcium channel blockers, antiepileptics and antipsychotics) targeted different subnetworks with association with schizophrenia. Known antipsychotics target dopamine, serotonin, adrenergic, and muscarinic acetylcholine receptors. The selective calcium channel blockers mainly target calcium channels, whereas antiepileptics target GABA receptors, glutamate receptors, sodium channels, calcium voltage-gated channels and nicotinic acetylcholine receptors. The top targets for antipsychotics (with highest association with schizophrenia) are *DRD2*, *CHRM4* and *HTR5A*; for antiepileptics, *CACNA1I* and *SCN9A*, and for selective calcium channel blockers, *CACNA1C* and *CACNB2*. The top genes in epilepsy pathways are *AKT3*, *GABBR2*, and *KCNQ2*, and the main target families are GABA receptors, glutamate receptors, potassium channels, sodium channels, and calcium voltage-gated channels (Fig. 4b).

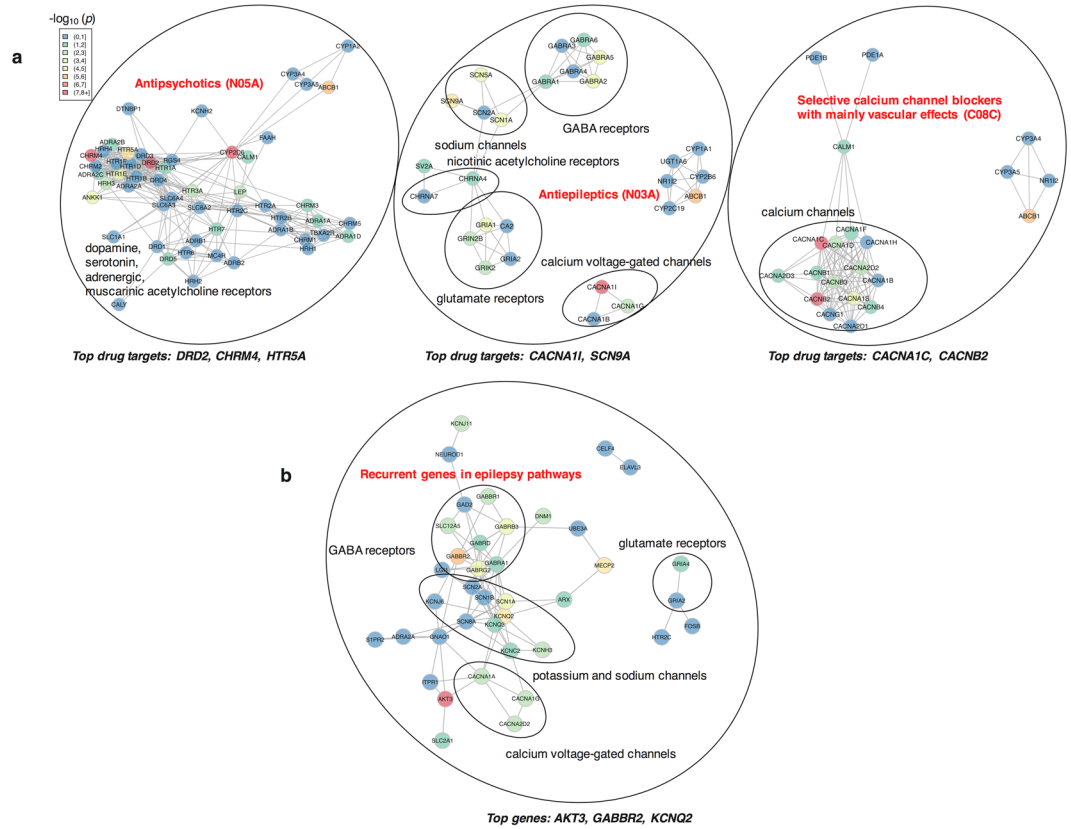

Figure 4. Protein-protein interaction networks. The interactions and interaction scores were obtained through the STRING[17] online platform. Vertices were placed on a plane using the Fruchterman-Reingold layout algorithm. Each node is colored by $-\log_{10}(p)$, which measures the degree of association of a gene with schizophrenia (SCZ-PGC2). **(a)** Protein-protein interaction networks for the three drug classes significant in SCZ-PGC2: only proteins targeted by at least 2 drugs within the class are shown. **(b)** Protein-protein interaction network in epilepsy pathways: only genes present in at least 10 epilepsy pathways from Open Targets are shown.

Discussion

We find that the targets of antipsychotics, the primary drug class used to treat schizophrenia, are enriched for association in current schizophrenia GWAS results. We also show that this enrichment increases with the number of schizophrenia cases included in the GWAS, the largest being SCZ-PGC2 (~35,000 cases). In addition, our results show significant enrichment for two other drug classes: selective calcium channel blockers and antiepileptics.

It is noteworthy that there is no evidence for a genetic correlation between schizophrenia and epilepsy as measured by linkage disequilibrium score (LDSC) regression[18]. However, our analyses reveal that epilepsy pathways and the targets of antiepileptics (with GABAergic and antiglutamatergic action) are enriched in schizophrenia. Some antiepileptics have also been investigated for treatment-resistant schizophrenia[19].

Voltage-gated channels have been widely studied in psychiatric disorders[20], and L-type calcium channels have been associated with schizophrenia in numerous studies[21]. Amongst top drugs targeting calcium channels, verapamil has been reported to be comparable to lithium for the treatment of mania[22]. Cinnarizine, which has atypical antipsychotic properties in animal models[23], is prescribed for vertigo because of its antihistamine properties and is also an antagonist of dopamine D2 receptors.

Nicotinic acetylcholine receptors show significant association in SCZ-PGC2. The *CHRNA3-CHRNA5-CHRNB4* gene cluster is strongly associated with schizophrenia; it consists of genes in high LD with each other and has been linked to nicotine dependence[24]. Some studies indicate that nicotine could have a positive effect on psychotic symptoms and cognitive function in schizophrenic patients[25]. These results are consistent with a recent study by Won *et al.* that also highlights the enrichment of acetylcholine receptor activity in schizophrenia[26]. Several drugs, such as varenicline and galantamine, target these receptors. Varenicline is a nicotinic agonist used for smoking cessation[27] while galantamine is an allosteric modulator of nicotinic receptors and an acetylcholinesterase inhibitor, and has been investigated for the treatment of cognitive impairment in schizophrenia[28].

Some 40 of the significant druggable genes are not located in schizophrenia GWAS loci. These include *GABBR2*, *NOS1* and *OPRD1*. Significant reduction in *GABBR2* protein expression has been reported in the lateral cerebellum of postmortem brains from schizophrenia, bipolar and major depressive disorder subjects in comparison to unaffected subjects[29]. Relevant hub genes (excluding MHC) with functional studies showing evidence of an effect on behaviour or the nervous system, are *CACNA1I*, *CHRM4*, *CHRNA3*, *CLCN3*, *CNTN4*, *NEK1*, and

OPRD1 (genes with highest normalized node degree), and *F2*, *MAPK3*, *POMC*, *ATP2A2*, *FGFR1*, *FURIN*, *EP300*, *CLU*, *MARK2* and *NOS1* (genes with highest normalized betweenness centrality).

Compounds targeting proteins encoded by *MCHR1* and *DPP4* might also be of particular interest. *MCHR1* antagonists include high affinity ligands such as ATC0175 or ATC0065, which exhibit antidepressive and anxiolytic effects in mouse and rat behavioral models[30]. *DPP4* inhibitors include gliptins such as dutogliptin and alogliptin, which are used to treat type 2 diabetes, and atorvastatin, which is prescribed due to its cholesterol-lowering properties[31]. Current antipsychotics can induce insulin resistance[32], and drugs which do not or would reverse these effects would be a welcome addition to the pharmacopoeia.

In summary, our workflow may be used identify new drug targets and repurposing opportunities, and visualise biological pathways. It is suitable for use as a filtering process in the first stages of drug discovery. We conclude that sufficiently powerful GWASs can be examined with increased confidence for drug target identification and repurposing opportunities across complex disorders, by investigating biological pathways, drug gene-sets and druggable genes. In disorders that have few known drug treatments, such as eating disorders and autism, verifying the signal of known drugs might not be possible, but once well-powered GWASs with multiple significant signals become available, this approach could still be effective to generate much needed therapeutic hypotheses.

Methods

Methods: Pathway analysis. The pathway analysis software MAGMA v. 1.06[33] was used to generate p-values for genes and gene-sets representing drugs, gene families, biological pathways and disease pathways. GWAS summary statistics are available as SNP p-values, which MAGMA combines to produce gene and gene-set p-values. We used a combined model with top and mean SNP associations to compute gene p-values. These gene p-values are converted to Z-values, which are used as the response variable in a regression model, solved using a generalized least squares approach accounting for linkage disequilibrium. Two types of regression analyses can be conducted: self-contained or competitive. The self-contained approach tests whether the pathway is associated with a trait of interest, whereas the competitive approach tests whether genes in the pathway are more strongly associated than genes outside the pathway. The self-contained approach is more powerful, but it is sensitive to the polygenic nature of observed GWAS statistics inflation and may lead to a higher Type I error[33,34]. Therefore, we used competitive p-values. In MAGMA, the competitive analysis corrects for gene size, density, minor allele count, and takes into account gene-gene correlations[33]. The SNP positions and frequencies were extracted from the European subset of 1000 genomes phase III v.5a[35] with genome assembly hg19. We used Ensembl release 75[36] for the gene positions. The gene window was set to 35 kb upstream and 10 kb downstream in MAGMA to include gene regulatory regions. We generated FDR-adjusted p-values or *q-values* for genes and gene-sets, using Benjamini and Hochberg's method to account for multiple testing[37]; we also provide the Benjamini and Yekutieli q-values[38] and Bonferroni-corrected p-values for all our results in Supplement 2.

Methods: Pathway maps. The top 50 biological pathways and top 50 drugs with ATC codes were used to produce two separate maps, using the p-values obtained in the pathway analysis step. Gene-sets were encoded by gene content (binary vectors) and a Tanimoto similarity matrix was generated. This matrix was used as input for the k-GTM algorithm[15,16] implemented in GTMapTool v1.0[39]. Five parameters need to be defined by the user: the square root of the number of sample points (k), the square root of the number of radial basis functions (RBF), the regularization coefficient (l), the RBF width factor (w) and the feature space dimension (D). We set $k =$ square root of the number of data points in the input kernel, $m =$ square root of k, $l = 1$ and $w = 1$ (default values). The feature space dimension D was estimated as the number of PCs explaining 99.5% of the variance in the input data. We used the same method to compute the number of independent tests and generate the Bonferroni-corrected p-values for pathways. The maps were colored by schizophrenia association in $-\log_{10}(p)$ units using the kriging algorithm implemented in the R package *gstat*[40].

Methods: Protein-protein interaction networks. Genes driving the association in pathway clusters or drug families were highlighted in protein-protein interaction networks. Protein-protein interaction scores were generated using the STRING v.10 online platform[17], which integrates information from genomic context predictions, high-throughput lab experiments, co-expression, automated textmining, and other databases. The Fruchterman-Reingold layout algorithm implemented in the R package *igraph*[41] was used to position the vertices on the graphs depending on the interaction score; each gene (node) was colored by its association with schizophrenia computed by MAGMA, in $-\log_{10}(p)$ units. The PPI network of top druggable genes was generated with the *STRINGdb*[17] R package, and *igraph* was used to compute the betweenness centrality and node degree for each gene (cf. Text S3 in Supplement 1).

Methods: Enrichment measure for groups of gene-sets. Instead of investigating individual gene-sets, we focused on *groups* of gene-sets. For example, a class of drugs can be represented by a group **S** of drugs (gene-sets). To determine whether **S** is significantly enriched, we can draw enrichment curves, widely used in virtual screening[42]. The curves display the percentage of hits found when decreasing the value of a scoring function. Here, the scoring function is the gene-set association with the trait of interest in $-\log_{10}(p)$ units, and the hits are the gene-sets. The area under this enrichment curve (AUC) provides a quantitative assessment of the enrichment of **S** in a GWAS and is computed using the trapezoidal approximation of an integral. The expected random result is $AUC = 50\%$ and the maximum value is $AUC = 100\%$.

The AUC significance was assessed using Wilcoxon-Mann-Whitney test (WMW), which tests whether the data distribution is the same within two different groups (e.g., gene-sets in **S** and not in **S**)[43] - also, the AUC can be directly calculated from the Wilcoxon-Mann-Whitney U statistic[44]. We used this enrichment measure to assess whether drugs in a set **S** were more associated with a disorder than others, while accounting for the fact that drug

gene-sets are diverse and noisy, due to an incomplete knowledge of targets, the presence of off-targets without any association with the disorder, and the fact that drugs may have different mechanisms of action within the same therapeutic class.

Materials: Schizophrenia GWAS summary statistics. In this paper, we used three GWASs conducted in 2011[13], 2013[14] and 2014[4] with increasing sample sizes (cf. Figure 3b and Table S1 in Supplement 1). The three studies were coined SCZ-PGC1, SCZ-PGC1+SWE and SCZ-PGC2, respectively. The three studies mainly contain individuals of European ancestry[4,13,14]; SCZ-PGC2 is the only study including the X chromosome and individuals with East Asian ancestry. Only SNPs present in the European subset of 1000 genomes phase 3 v.5a[35] with minor allele frequency (MAF) $\geq$ 1% were kept. The genomic inflation factor as well as the LD score intercept were computed for each set using the LDSC software v.1.0.0[45]. All p-values were subsequently corrected using the LD score intercept - a score based on linkage disequilibrium that should provide a better way to control for inflation than the genomic inflation factor[46]. Only the 1,123,234 SNPs shared among SCZ-PGC1, SCZ-PGC1+SWE and SCZ-PGC2 were considered when comparing the three studies. The latest and most powerful GWAS (SCZ-PGC2) was used to investigate the enrichment of drug classes, drug gene-sets, and biological pathways.

Materials: Druggable genome. We used the 4479 genes in the "druggable genome" defined by Finan *et al.*[47], divided into 3 Tiers based on druggability levels: Tier 1 contains genes encoding targets of approved or clinical trial drugs, Tier 2 genes encoding targets with high sequence similarity to Tier 1 proteins or targeted by small drug-like molecules, and Tier 3 contains genes encoding secreted and extracellular proteins, genes belonging to the main druggable gene families, and genes encoding proteins with more restricted similarity to Tier 1 targets. In the pathway-wise analyses, genes were used whether or not they were present in the druggable genome, but only druggable genes outside the MHC were investigated to prioritize druggable targets.

Materials: Drug gene-sets. Drug-gene interactions are mainly derived from drug-target activity profiles. The data was drawn from two sources: the Drug-Gene Interaction database DGIdb v.2[48], and the Psychoactive Drug Screening Database K_i DB[49] downloaded in June 2016. DGIdb is a new resource that integrates drug-gene interactions from 15 databases, amongst which DrugBank and ChEMBL; the data is directly available as drug-gene pairs and genes are identified by their HGNC (HUGO Gene Nomenclature Committee) names[50]. K_i DB provides K_i values for drug/target pairs and is particularly relevant for psychoactive drugs. More details on the filtering procedure can be found in Text S1 in Supplement 1. Gene-sets were produced by merging both DGIdb and K_i DB drug/gene data and by converting HGNC names to Ensembl release 75[36] identifiers. The number of unique gene-sets was 3939 at the end of the filtering process, 3913 with variants in SCZ-PGC2 (2586 independent gene-sets), out of which 1026 were mapped to at least one ATC code. We annotated groups of drugs using ATC classes, listed in Table S9 in Supplement 2 and containing at least 10 drugs. The drug set used to check the enrichment of antipsychotics in schizophrenia GWASs is the set of drugs with ATC code N05A - all schizophrenia drugs belong to this class (cf. Table S2 in Supplement 1 for the list of prescription drugs in the UK).

Materials: Biological pathways. We refer to our ensemble of gene ontology pathways, canonical pathways, disease pathways, and gene families as "biological pathways". Canonical (CP) and Gene Ontology (GO) gene-sets were extracted from MSigDB v5.2[51]. MSigDB is a regularly updated resource gathering pathways and ontologies from the main online databases. CP sets were curated from: BioCarta, KEGG, Matrisome, Pathway Interaction Database, Reactome, Sigma Aldrich, Signaling Gateway, Signal Transduction KE and SuperArray. These "pathways" provide a practical way to investigate the function of a subnetwork without accounting for the complexity of biological networks. Disease pathways were extracted from the Open Targets platform[52] in January 2017 and gene families were identified using information provided on the HGNC website. The total number of biological pathways was 13,572 (9408 independent pathways).

Data and materials availability. All data used in this paper are freely available online (cf. references and supplementary materials).

References

1. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
2. Lichtenstein, P. *et al*. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).
3. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. USA*. **58**, 199–205 (1967).
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
5. Lencz, T. & Malhotra, A. K. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. *Mol. Psychiatry* **20**, 820–826 (2015).
6. Nelson, M. R. *et al*. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).
7. Sanseau, P. *et al*. Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* **30**, 317–320 (2012).
8. Wang, Z.-Y. & Zhang, H.-Y. Rational drug repositioning by medical genetics. *Nat. Biotechnol.* **31**, 1080–1082 (2013).
9. Breen, G. *et al*. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19**, 1392–1396 (2016).
10. Ruderfer, D. M. *et al*. Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach. *Lancet Psychiatry* **3**, 350–357 (2016).
11. de Jong, S., Vidler, L. R., Mokrab, Y., Collier, D. A. & Breen, G. Gene-set analysis based on the pharmacological profiles of drugs to identify repurposing opportunities in schizophrenia. *J. Psychopharmacol.* **30**, 826–830 (2016).
12. Chang, S., Fang, K., Zhang, K. & Wang, J. Network-Based Analysis of Schizophrenia Genome-Wide Association Data to Detect the Joint Functional Association Signals. *PLoS One* **10**, e0133404 (2015).
13. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet*. **43**, 969–976 (2011).

14. Ripke, S. *et al*. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
15. Bishop, C. M., Svensén, M. & Williams, C. K. I. GTM: The Generative Topographic Mapping. *Neural Comput.* **10**, 215–234 (1998).
16. Olier, I., Vellido, A. & Giraldo, J. Kernel generative topographic mapping. In *ESANN 2010, 18th European Symposium on Artificial Neural Networks, Bruges, Belgium, April 28–30, 2010, Proceedings* (2010).
17. Szklarczyk, D. *et al*. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–52 (2015).
18. Anttila, V. *et al*. Analysis of shared heritability in common disorders of the brain. *bioRxiv* 048991. https://doi.org/10.1101/048991 (2016).
19. Hosák, L. & Libiger, J. Antiepileptic drugs in schizophrenia: a review. *Eur. Psychiatry* **17**, 371–378 (2002).
20. Imbrici, P., Camerino, D. C. & Tricarico, D. Major channels involved in neuropsychiatric disorders and therapeutic perspectives. *Front. Genet.* **4**, 76 (2013).
21. Berger, S. M. & Bartsch, D. The role of L-type voltage-gated calcium channels Cav1.2 and Cav1.3 in normal and pathological brain function. *Cell Tissue Res.* **357**, 463–476 (2014).
22. Dubovsky, S. L. & Buzan, R. The Role of Calcium Channel Blockers in the Treatment of Psychiatric Disorders. *CNS Drugs* **4**, 47–57 (1995).
23. Dall'Igna, O. P., Tort, A. B. L., Souza, D. O. & Lara, D. R. Cinnarizine has an atypical antipsychotic profile in animal models of psychosis. *J. Psychopharmacol.* **19**, 342–346 (2005).
24. Greenbaum, L., Rigbi, A., Teltsh, O. & Lerer, B. Role of genetic variants in the CHRNA5-CHRNA3-CHRNB4 cluster in nicotine dependence risk: importance of gene-environment interplay. *Mol. Psychiatry* **14**, 828–830 (2009).
25. Chen, J. *et al*. Genetic Relationship between Schizophrenia and Nicotine Dependence. *Sci. Rep.* **6**, 25671 (2016).
26. Won, H. *et al*. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. https://doi.org/10.1038/nature19847 (2016).
27. Evins, A. E. *et al*. Maintenance treatment with varenicline for smoking cessation in patients with schizophrenia and bipolar disorder: a randomized clinical trial. *JAMA* **311**, 145–154 (2014).
28. Buchanan, R. W. *et al*. Galantamine for the treatment of cognitive impairments in people with schizophrenia. *Am. J. Psychiatry* **165**, 82–89 (2008).
29. Fatemi, S. H., Folsom, T. D. & Thuras, P. D. Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. *Schizophr. Res.* **128**, 37–43 (2011).
30. Chaki, S. Anxiolytic- and Antidepressant-Like Profile of ATC0065 and ATC0175: Nonpeptidic and Orally Active Melanin-Concentrating Hormone Receptor 1 Antagonists. *J. Pharmacol. Exp. Ther.* **313**, 831–839 (2004).
31. Taldone, T., Zito, S. W. & Talele, T. T. Inhibition of dipeptidyl peptidase-IV (DPP-IV) by atorvastatin. *Bioorg. Med. Chem. Lett.* **18**, 479–484 (2008).
32. Teff, K. L. *et al*. Antipsychotic-induced insulin resistance and postprandial hormonal dysregulation independent of weight gain or psychiatric disease. *Diabetes* **62**, 3232–3240 (2013).
33. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
34. Maciejewski, H. Competitive and self-contained gene set analysis methods applied for class prediction. in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on* 55–61 (2011).
35. 1000 Genomes Project Consortium. *et al*. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
36. Yates, A. *et al*. Ensembl 2016. *Nucleic Acids Res.* **44**, D710–6 (2016).
37. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
38. Benjamini, Y. & Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
39. Gaspar, H. A., Baskin, I. I., Marcou, G., Horvath, D. & Varnek, A. GTM-Based QSAR Models and Their Applicability Domains. *Mol. Inform.* **34**, 348–356 (2015).
40. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* **30**, 683–691 (2004).
41. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* **1–9**, 2006 (**1695**).
42. Varin, T., Schuffenhauer, A., Ertl, P. & Renner, S. Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data. *J. Chem. Inf. Model.* **51**, 1528–1538 (2011).
43. Wilcoxon, F. Individual Comparisons by Ranking Methods. In *Breakthroughs in Statistics* (eds Kotz, S. & Johnson, N. L.) 196–202 (Springer New York, 1992).
44. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
45. Brendan Bulik-Sullivan, H. F. *LD Score Regression (LDSC)*. (Broad Institute of MIT and Harvard / MIT Department of Mathematics, 2015).
46. Bulik-Sullivan, B. K. *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
47. Finan, C. *et al*. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9** (2017).
48. Wagner, A. H. *et al*. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res.* **44**, D1036–44 (2016).
49. Roth, B. L., Lopez, E., Patel, S. & Kroeze, W. K. The Multiplicity of Serotonin Receptors: Uselessly Diverse Molecules or an Embarrassment of Riches? *Neuroscientist* **6**, 252–262 (2000).
50. Gray, K. A., Yates, B., Seal, R. L., Wright, M. W. & Bruford, E. A. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.* **43**, D1079–85 (2015).
51. Subramanian, A. *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
52. Koscielny, G. *et al*. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* **45**, D985–D994 (2017).

Acknowledgements

K_i determinations were generously provided by the National Institute of Mental Health's Psychoactive Drug Screening Program, Contract # HHSN-271-2013-00017-C (NIMH PDSP). The NIMH PDSP is Directed by Bryan L. Roth MD, PhD at the University of North Carolina at Chapel Hill and Project Officer Jamie Driscoll at NIMH, Bethesda MD, USA. HG and GB acknowledge funding from the US National Institute of Mental Health (PGC3: U01 MH109528). This work was also supported in part by the NIHR Maudsley Biomedical Research Centre ('BRC') hosted at King's College London and South London and Maudsley NHS Foundation Trust, and funded by the National Institute for Health Research under its Biomedical Research Centres funding initiative. The views expressed are those of the authors and not necessarily those of the BRC, the NHS, the NIHR or the Department

of Health or King's College London. We gratefully acknowledge capital and computing equipment funding from the Maudsley Charity (Grant Reference 980) and Guy's and St Thomas's Charity (Grant Reference STR130505).

Author Contributions

H.G. produced the results and introduced the pathway visualisation methodology, G.B. provided advice and guidelines, both H.G. and G.B. contributed to the methodology and the writing of the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-12325-3.

Competing Interests: G.B. reports consultancy and speaker fees from Eli Lilly and Illumina and grant funding from Eli Lilly.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.